Fall Creek Falls – Session IIa

Key Challenges presented by next generation hardware systems
Session Agenda

- **Doug Burger - Texas**
  - Termites, Chainsaws, and Bulldozers: The Case for Composable Processors

- **Dimitris Nikolopoulos - Virginia Tech**
  - System Software Challenges and Opportunities on Asymmetric Multi-core Systems

- **Volodymyr Kindratenko - NCSA/UIUC**
  - High-Performance Computing on FPGAs: Challenges and Opportunities

- **John Levesque - Cray**
  - Scaling Beyond Commodity

- **Sid Chatterjee - IBM**
  - Breaking Down Walls

- **Panel**
  - All speakers
Heterogeneous Architectures
Coming Soon to a System Near You

Jeffrey Vetter
ORNL/Georgia Tech
Best of times, worst of times...

- **Years of Prosperity**
  - Increasing large-scale parallelism
  - Increasing number of transistors
  - Increasing clock speed
  - Stable programming models and languages

- **‘New’ Constraints on architectures**
  - Power
  - Heat / thermal envelope
  - Signaling
  - Packaging
  - Instruction level parallelism
  - Memory, I/O, interconnect latency and bandwidth
  - Market trends favor ‘good enough’ computing – *Economist*
Architectures – Multicore

- Core counts (peak flops) are easy to increase

Other resources are the challenge!
Architectures – Heterogeneity, Specialization

Architectures target specific workloads: games, graphics, business, encryption, media, etc.
Case Study of Molecular Dynamics across Various Architectures

- **Bio-molecular simulation techniques: Molecular Dynamics**
- **MD simulations**
  - $N^2$ complexity (reduced by cutoff)
  - $N$ order of 10K to over 1000K atoms
  - Unit of simulation time: femto seconds
  - Desirable simulation: with 100K atoms over $10^{-3}$ and $10^{-6}$ seconds

- **Current capabilities**
  - A few hundred pico-seconds per simulation day on a multi-core desktop system
  - With AMBER on MPP systems: few hundred pico-seconds on up to 128 processors
  - With LAMMPS on MPP systems: couple of nano-seconds on ~10K processors

- **Focused on accelerating non-bonded calculations**

Contemporary Heterogeneous Architectures

Cell Chip Block Diagram

- SXU
- LS
- SMF
- EIB (up to 96 Bytes/cycle)
- L2
- PPE
- MIC
- BIC
- FlexIO™

SPU SPE

Sub-problem A

Sub-problem B

Sub-problem C

Sub-problem D

Unused streams

Programs running in parallel

Concurrent threads of computation

Hardware streams (128)

Instruction Ready Pool

Pipeline of executing instructions

Vertex Shader Units

Pixel Pipelines

ROP Engine

Memory Partition

Memory Partition

Memory Partition

Memory Partition

Dual XDR™

Serial Code

L1
Cell

Runtime (sec)

<table>
<thead>
<tr>
<th>Task</th>
<th>0.00</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>original</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>replace &quot;if&quot; with &quot;copysign&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMD unit cell reflection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMD direction vector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMD length calculation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIMD acceleration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Runtime (sec)

1 SPE  | 8 SPEs  | 1 SPE  | 8 SPEs |
Respawn every time step | 4.701 sec | 0.181 sec | 0.816 sec | 0.181 sec |
Launch only first time step | 0.925 sec | 0.816 sec | 0.181 sec | 4.701 sec |

Number of Atoms | 2048

<table>
<thead>
<tr>
<th>Processor</th>
<th>Runtime (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opteron</td>
<td>0.925 sec</td>
</tr>
<tr>
<td>Cell, 1 SPE</td>
<td>0.816 sec</td>
</tr>
<tr>
<td>Cell, 8 SPEs</td>
<td>0.181 sec</td>
</tr>
<tr>
<td>Cell, PPE only</td>
<td>4.701 sec</td>
</tr>
</tbody>
</table>
GPU

- NVIDIA GeForce 7900GTX GPU versus a 2.2GHz Opteron
- 6x speedup including data transfer overheads
MTA-2

- **Realities**
  - 200 MHz vs. 2.2 GHz
- **Opteron ~2x faster**
- **Compared scaling with increasing number of atoms**
- **Opteron runtime increases faster rate**
MD Conclusions

- **Comparison:**
  - Productivity vs. performance
  - Cost effectiveness and availability

- **Source line of code comparison (SLOC)**
  - GPU highest
  - MTA-II lowest

- **Performance comparison (2048 atoms):**
  - 6x on GPU, increase with number of atoms
  - 5x on Cell
  - 0.5x on MTA-II

- **Can we sustain these for:**
  - Full code acceleration
  - Double-precision calculations

- **Is performance gap going to increase/decrease with multi-core systems?**
Key Experiences (and Challenges)

- No common software stack across platforms
  - OS, runtime, virtualization different for each system
  - Lhotse software stack

- Performance projections
  - Uncertainty in architecture/application mapping
  - Modeling Assertions for symbolic application model

- Performance instabilities
  - Small changes in source code can cause big changes in performance
  - Improved development environments, compilers, and libraries


Acknowledgements

- Thanks to contributors, sponsors
  - Experimental Computing Lab @ ORNL

- More information

- Acknowledgements
  - This research was sponsored by the Office of Mathematical, Information, and Computational Sciences, Office of Science, U.S. Department of Energy under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.